DOCUMENT RESUME

ED 127 893 HE 008 208

AUTHOR Pezzullo, Thomas R.; And Others
TITLE Student Evaluation of Instruction: A Validity Analysis.
PUB DATE May 76
NOTE 10p.; Paper presented at the annual meeting of the Association for Institutional Research (Los Angeles, California, May 197[illegible])
AVAILABLE FROM Curriculum Research and Development Center, University of Rhode Island, Kingston, Rhode Island 02881

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Demography; *Evaluation; *Evaluation Methods; Factor Analysis; *Higher Education; Student Characteristics; *Student Evaluation of Teacher Performance; *Student Opinion; *Validity

ABSTRACT

The use of student ratings of college instructors has steadily increased, with an attendant increase in the use of these student ratings in decision-making related to merit increases, promotion, tenure, and institutional severance. While a substantial body of research on student rating of instruction exists, the ambiguous or actually conflicting results of several of these studies has also led to concern by many professionals about the functional utility of student ratings. Using a sample of nearly 2,000 courses offered at the University, comparisons were made of: (1) correlations between Global Instructor Rating (GRI) and static course and student characteristics; and (2) predictor variables, order, regressions of student and course characteristics on GRI. The study collected one of the largest and most comprehensive sets of data on the subject of student evaluations of teaching. On the basis of the analysis it appears that, at a minimum, only a rather small portion of the total variance in instructor ratings can be attributed to demographic characteristics over which they have little control. (Author/KE)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127893

CURRICULUM RESEARCH AND DEVELOPMENT CENTER

UNIVERSITY OF RHODE ISLAND, KINGSTON, RHODE ISLAND 02881

STUDENT EVALUATION OF INSTRUCTION: A VALIDITY ANALYSIS

Thomas R. Pezzullo, University College

John V. Long, Curriculum Research and Development Center

Roy M. Ageloff, Management Science Department

University of Rhode Island

Presented at: The Annual Meeting of the Association for Institutional Research, Los Angeles, California, May, 1976

HE 008 208

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BACKGROUND

In the past ten years the use of student ratings of instructors on our college campuses has steadily increased, with an attendant increase in the use of these student ratings in decision making related to merit increases, promotion, tenure, and institutional severance. A survey of 410 college deans found that in the period from 1966 to 1973 the data source used by the deans to evaluate teaching which showed the greatest increase in frequency of use was systematic student ratings of instructors (Seldin, 1974). With the recent AAUP Statement on Teaching Evaluation (1974), which asserts that "student perceptions are a prime source of information from those who must be affected if learning is to take place. Student responses can provide continuing insights into a number of dimensions of a teachers efforts ..." (p. 169), it can be assumed that the use of systematically collected student perceptions will become even more widespread in professional and instructional evaluation.

While a rather substantial body of research on student rating of instruction currently exists (see Trent and Cohen, 1973; Costin, Greenough, and Menges, 1971; and Centra and Creech, 1976 for reviews), the ambiguity and/or actually conflicting results of several of these studies has also led to concern by many professionals about the functional utility of student ratings. Gage (1961), for example, stated that "teachers should not be penalized because of conditions over which they have no control such as level of the course, size of the class, and whether the course is elective or required." (p. 17). Because he felt these conditions affected student ratings he urged that such ratings not be used for purposes of promotion of institutional severance. Other, more recent

statements, (Kerlinger, 1971; Peck, 1971; Anthony and Lewis, 1972) have supported Gage's positions and sparked continued debate over the use of student ratings of instructors for institutional decision-making.

Centra and Creech (1976, p. 11) report that theirs and most other prior investigations have reached the conclusion that students with better grade-point averages do not necessarily rate teachers more favorably, although students who expected a lower grade than their own grade point average tended to rate their teacher as less effective. They call this a "modest source of bias in an overall rating of teacher performance." (Ibid, p. 13). Centra and Creech also conclude that course-level and student-level produced little difference in ratings (Ibid, p. 15).

They also indicate that in the analysis of over 8,000 instructors, faculty rank produced no significant differences in rating, except that teaching assistants received lower ratings than the four regular faculty ranks, (Ibid, p. 20).

With respect to course type Centra and Creech (Ibid, p. 30) also concluded that courses conducted in the strict lecture mode received the lowest ratings.

On class size they deduced that while some studies have reported no relationship, others show a slight negative trend and their own observations show considerable variability from size to size with the smallest and the largest classes receiving the generally higher ratings.

There seems to be considerable variation in results and in approach in the analysis of instructional evaluation data. This study was, therefore, undertaken to determine whether significant proportions of the variance in the students' course and faculty ratings are attributable to student demographic characteristics or static course or faculty

characteristics beyond the control of the instructor. Our interest was sparked by Gage's claim that these factors unduly affect the student's attitude toward the course and instructor. We, like others, are concerned that the Student Instructional Report (SIR), our evaluation questionnaire (ETS, 1971) measured behavior-specific facets of instructional performance and are not "unduly affected" by variables which the instructor cannot control.

METHOD AND RESULTS

The principal method for estimating variance explained by static course variables and predictability of faculty ratings was stepped multiple regression. The sample employed was all of the nearly 2,000 courses (37,000 students) offered at the University in one semester, which guarantees both a substantial sample size and a comprehensive range of course types, level, size and academic field. For each class the ratings for all students were pooled, and the mean scores represented the element of data in the regression analysis.

Static course variables available for the regression analysis were as follows: (1) Expected grade in course, (2) Class size, (3) Student ability (self-reported prior grades), (4) Required vs. elective course, (5) Rank of instructor, (6) Instructor's number of years of teaching experience, (7) Instructor's teaching load, (8) Course type (lecture, discussion, lab, etc.,), (9) Course level (lower division, upper, graduate).

The criterion variable was the score on the final global item in the SIR questionnaire: "Compared with other instructors you have had, how effective has the instructor been in this course?", with ratings from "excellent"-(5) to "poor"-(1). This criterion variable will be referred to as the global instructor rating (GIR). An array of correlations of

these static course variables with the GIR criterion appears below in Table 1.

Table 1

Correlations between Global Instructor Rating (GIR) and Static Course and Student Characteristics

	Global Instructor Rating (GIR)
(1) Expected Grade	.20*
(2) Class Size	.20*
(3) Student Ability	.15*
(4) Req./Elect.	.03
(5) Rank	.03
(6) Teaching Experience	.05
(7) Teaching Load	.05
(8) Course Type	.05
(9) Course Level	.15*

*p .01, N=1930

Inspection of Table 1 reveals that the criterion measure, GIR, is significantly correlated with expected grade, class size, student ability (grade point average) and course level. To examine the predictive power of these static course variables, and to examine their combined effect upon the overall rating of the instructor, stepwise multiple regression was conducted. Table 2 presents the results of the stepwise regression.

Table 2

Predictor Variables, Order, R and R^2 for Regressions of Student and Course Characteristics on Global Instructor Rating

Var #	Variable Entered	R	R^2	Increase
(2)	Class Size	.204	.042	-
(1)	Expected Grade	.264	.070	.028
(3)	Student Ability	.275	.076	.006
(4)	Required/Elect.	.284	.080	.004
(9)	Course Level	.287	.082	.002
(7)	Teaching Load	.290	.084	.002
(6)	Teaching Experience	.291	.085	.001
(8)	Course Type	.291	.085	.000
(5)	Teacher Rank	.291	.085	.000

As Table 2 reveals, the combined predictive power of all of the static course characteristics is low (R= .291). Nonetheless, a statistically significant proportion of variance (8 1/2%) can be explained by student or course characteristics beyond the control of the instructor. The regression indicates that among all the variables studied prediction rests most heavily upon class size and grade expectation. While one might ordinarily be disappointed at the low degree of predictability represented by these data, we are relieved that so little of student evaluation of the instructor can be explained by static course faculty and student characteristics.

CONCLUSIONS

Since the evaluation of courses and instruction is a delicate area of controversy, it is important to determine that a minimal share of the variance in student response is contingent upon static course and/or demographic characteristics in order to interpret the ratings with some degree of confidence. Were they to have been highly predictable the results of much student opinion-based evaluation would have to be qualified by each of the significant, related demographic and course characteristics. This study suggests that they may be interpreted in a more straightforward manner. Reduced predictive power implies greater independence from non-evaluative characteristics which are outside the control of the instructor and enables more reliable input to both instructional and administrative decision-making. The instructor then is encouraged to take this feedback seriously as is the committee of peers, chairpersons or deans who review these data for purposes of evaluation.

There are, of course, numerous other sources of data on instructional evaluation: peer review, self-evaluation, chairman and dean's personal review, and even "outcome" evaluation, i.e., student performance on standardized or criterion-specific tests, (this latter area being both the most primitive and most professionally unsettling). Nonetheless, a report on instructional evaluation would be incomplete without at least acknowledging them. This project sought to focus on only one source of data -- one which has grown in popularity since the mid-sixties and will probably continue to grow.

The authors conclude that while we may not have permanently laid to rest the notion that student evaluations of teaching are unreliable

and of limited validity, we have gathered one of the largest and most comprehensive sets of data on the subject and on the basis of the analysis it appears that, at a minimum, this study has illustrated that only a rather small portion of the total variance in instructor ratings can be attributed to demographic characteristics over which they have little control.

REFERENCES

AAUP Statement on Teaching Evaluation, AAUP Bulletin, Vol. 60, No. 2, June, 1974.

Aleamoni, L. and Graham, M.: The Relationship Between CBQ Ratings and Instructor's Rank, Class Size and Course Level; Journal of Educational Measurement: Vol. 11, No. 2, Pp. 189-201, Fall, 1974.

Anthony W. and Lewis, J. Learning effectiveness: The measure of effective teaching. American Association of Collegiate Schools of Business Bulletin, Vol. 9, October 1972.

Centra, J. and Creech, F.: The Relationship Between Student, Teacher and Course Characteristics and Student Ratings of Teacher Effectiveness, Educational Testing Service, March, 1976.

Costin, F. Greenough, W., and Menges, R.; Student ratings of college teaching: reliability, validity and usefulness; Review of Educational Research, Vol. 41, No. 5, Pp. 511-533, December, 1971.

Gage, N. L. The appraisal of college teaching. Journal of Higher Education, Vol. 32, Pp. 17-22, 1961.

Granzin, K. and Painter, J., A New Explanation for Students' Course Evaluation Tendencies; American Educational Research Journal, Vol. 10, No. 2, Pp. 115-124, Spring, 1973.

Greenwood, G., Bridges C., Ware, W. and MacLean, Jr.; Student evaluation of college teaching behaviors; Journal of Educational Measurements; Vol. 11, No. 2; Pp. 141-143, Summer, 1974.

Kerlinger, F. Student evaluation of university professors. School and Society, Vol. 99, Pp: 353-356, October, 1971.

McKeachie, N. J., Lin. Y., and Mann, W., Student ratings of teacher effectiveness: validity studies, American Educational Research Journal, Vol. 8, No. 4, 1971, Pp. 435-445.

Peck, R. Can students evaluate their education? Education Digest, Vol. 36, Pp. 29-31, 1971.

Seldin, P. How Colleges Evaluate Professors, Blythe-Pennington, New York, 1975.

Student Instructional Report, Education Testing Service, Princeton, New Jersey, 1971.

Trent, J. and Cohen, A. Research on teaching in higher education in The Handbook of Research on Teaching, R. Travers, Ed., Rand McNally and Co., Chicago, 1973, Pp. 997-1071.